A new Reliability-Based Data-Driven approach to Simulation-Based Models

Jacobo Ayensa Jiménez*, Mohamed Hamdy Doweidar* and Manuel Doblaré Castellano*
Mechanical Engineering Department, University of Zaragoza, Spain;
Aragón Institute of Engineering Research (I3A), University of Zaragoza, Spain;
Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Spain.
Campus Río Ebro, Edificio I+D, Mariano Esquillor S. N.
Email:jacoboaj@unizar.es; mohamed@unizar.es; mdoblare@unizar.es

***Abstract*—Data Science has burst into simulation-based engineering sciences with an impressive impulse. However, data are never uncertainty-free and a suitable approach is needed to face data measurement errors and their intrinsic randomness in problems with well-established physical constraints. We face this problem by hybridizing a standard mathematical modeling approach with a new data-driven solver accounting for the phenomenological part of the problem and able to handle the uncertainty of input data in an intelligent way. The reliability-based data-driven procedure performance is evaluated in a simple but illustrative unidimensional problem showing, in contrast with other data-driven solvers, better convergence, higher accuracy, clearer interpretation and major flexibility.**

I. Introduction

Despite the wide application of Data Science in areas such as marketing and e-commerce [1], social sciences [2] or healthcare [3], there are other fields where very little has been done. An example are the disciplines where physical models and the corresponding mathematical and numerical simulation tools are well established like Computational Physics, Computational Chemistry or Computational Engineering (simulation-based engineering and sciences -SBES-). A straightforward application of these techniques is dynamic data-driven systems (DDDAS) [4], in which the idea is providing both predictive and learning capabilities to the control system from data acquired from a sufficient set of sensors. This paradigm was settled down by Kalman [5] in the sixties with his ground-breaking filter and is still nowadays a hot topic of research opening up a huge range of possibilities [6].

Unlike these approaches, based on the direct treatment of available data, SBES incorporates, in addition to data, some *a priori* characteristic physical knowledge of the analyzed system. At this point, it is crucial to distinguish between two kinds of knowledge. On one hand, physical general principles, such as conservation and thermodynamic laws that are universally accepted as able to describe the underlying universe structure. On the other hand, phenomenological models, such as macroscopic material constitutive relations. Even if it would be possible to derive the real mechanistic constitutive relations also from first physical principles, the overwhelming number of degrees of freedom involved in the relevant spatial-temporal scales needed for real applications makes this possibility intractable [7]. *Data Analytics* techniques would be very useful in SBES to extrapolate the phenomenological submodel, but now constrained with the mathematical expression of first principles.

One idea in this direction was introduced by T. Sussman [8] for hyperelastic isotropic materials using splines interpolation in what is now known as "What You Prescribe Is What You Get" (WYPWYG) philosophy. Recently, F. Chinesta and coworkers defined a strategy for data-driven Computational Mechanics [9], combining Manifold Learning techniques and a (possibly optimized) directional search strategy inspired in the LaTin method [10]. M. Ortiz and his group presented a material model-free method based on the minimization of the distance between the searched solution and a set of experimental data, using a proper energy norm, while remaining in the equilibrium manifold, or equivalently, a well-posed penalty approach [11]. None of these works take into consideration the inherent inaccuracy of the data. Here, a new family of methods, called reliability-based data-driven solvers are defined. Data-driven solver methodology naturally allows incorporating reliability along the statement of the modeling. With this, simulations become sensitive to measurements precision and incorporate uncertainty considerations. An easy but illustrative one-dimensional problem is used to compare results and to show improvements using this methodology, emphasizing the added value with respect to existing methods.

II. Methodology

Data-driven solvers may be seen as iterative solvers searching for the intersection of an empirical (data based) manifold and a physical manifold. The first one is in many practical applications experimentally based and has, therefore, a discrete nature. The second is usually established in terms of sound laws particular to the problem in hands, but otherwise, derived from first principles universally accepted as the basis of Physics. For the sake of simplicity, let us consider the elastic three-dimensional problem. In that case, the physical manifold is the set of states that verify global and local equilibrium (i.e. conservation of linear and angular momentum), that in the static case (negligible inertial effects) is written in differential form as $\boldsymbol{\nabla} \cdot \boldsymbol{\sigma} = \mathbf{0}$, where $\boldsymbol{\sigma}$ is the stress tensor. This equation is usually approximated and solved in a discrete form using numerical methods like Finite Elements (FEM). After a convenient discretization, we can state $\mathbf{B}\mathbf{y} = \mathbf{0}$, where $\mathbf{y}$ is a finite dimensional vector containing the full stress tensor field information related to a given discretization (for FEM, this vector contains the components of the stress tensor for all the integration points) and $\mathbf{B}$ is a matrix encoding the geometry and connectivity of the domain.

The empirical manifold is defined via a set, $\mathcal{E} = \{(\mathbf{x}^j; \mathbf{y}^j)\}_{j=1,\cdots,m}$, of data points, resulting from experimental measurements. The set $\mathcal{E}$ may be seen as a representation of the underlying material behavior in the following asymptotic sense: (i) if $\mathcal{E}$ approximates a mathematical manifold and (ii) the uncertainty of each point approximates to zero.

A. *Problem formulation*

We postulate that a model-free engineering problem may be defined in terms of state variables (X, Y) that are related through a latent and unknown relationship $F(X, Y) = 0$. Returning to the discretized elastic problem, $\mathbf{x}_i$ is the vector containing all strain components, ε_{kl}, at the point i and $\mathbf{y}$ the vector containing all stress components. σ_{kl}, at the point i. It is now necessary to define a metric distance in the state space for $\mathbf{x}_i$ and $\mathbf{y}_i$. As we are considering engineering problems, we have physical constraints. For the sake of simplicity, but without any conceptual limitation, we shall consider linear constraints only, so they can be written as $\mathbf{A}\mathbf{x} = \mathbf{a}$ and $\mathbf{C}\mathbf{y} = \mathbf{c}$.

At each point i, we have a trial set $\mathcal{E}_i$ that may be thought as the result of experimental tests. We then define a local penalty function for each point i as:

$$F_i(\mathbf{x}_i, \mathbf{y}_i) = \min_{(\mathbf{x}', \mathbf{y}') \in \mathcal{E}_i} \{d_{x,i}(\mathbf{x}_i, \mathbf{x}') + d_{y,i}(\mathbf{y}_i, \mathbf{y}')\} \quad (1)$$

Therefore, a data-driven simulation-based engineering problem is defined by a constrained optimization problem where two steps are required:

- Local search of a minimum of the penalty function F_i for each element i using the nearest neighbor algorithm. This search looks for the most representative datum in the empirical discrete manifold.
- Global resolution of a linear system constraining the searched points to remain on the physical manifold.

B. *Reliability-based data-driven solver*

Now, each of the pairs $U^j = (X^j, Y^j)$ is considered to have random nature. Returning to the discrete case, $U = (X, Y) = (\mathbf{X}_i, \mathbf{Y}_i)_{i=1,\cdots,N}$, where $\mathbf{X}_i$, and $\mathbf{Y}_i$ are now random vectors whose dimension, n, is the size of the state vector and, as before, $\mathcal{N} = N \times n$ is the number of scalar state variables. We may, therefore, define a stochastic analogous problem to the deterministic one.

With this formulation, the solution candidate $\mathbf{u} = (\mathbf{x}, \mathbf{y})^T$ is not random, while $F(\mathbf{x}, \mathbf{y}|\mathcal{E}) = F(\mathbf{u}, \mathcal{E})$ is a random variable due to the random nature of $\mathcal{E}$, which can be characterized by means of expected value $\boldsymbol{\mu}$ and the variance-covariance matrix $\boldsymbol{\Sigma}$. A crucial point is to select a suitable norm for this stochastic approach of the problem. A very recommended one is Mahalanobis distance [12], equivalent to choose $\mathbf{M} = 2(\boldsymbol{\Sigma})^{-1}$ as metric matrix.

Under normality conditions, $D^2_{\mathcal{E}}$ follows a non-central chi-squared distribution with $n = 2\mathcal{N}$ degrees of freedom and non-centrality parameter $\lambda = (\mathbf{u} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{u} - \boldsymbol{\mu})$. In particular, expected value writes $\mu(D^2_{\mathcal{E}}) = 2\mathcal{N} + d^2_{\mathcal{E}}$ and variance $\sigma^2(D^2_{\mathcal{E}}) = 2(2\mathcal{N} + 2d^2_{\mathcal{E}})$.

III. Results

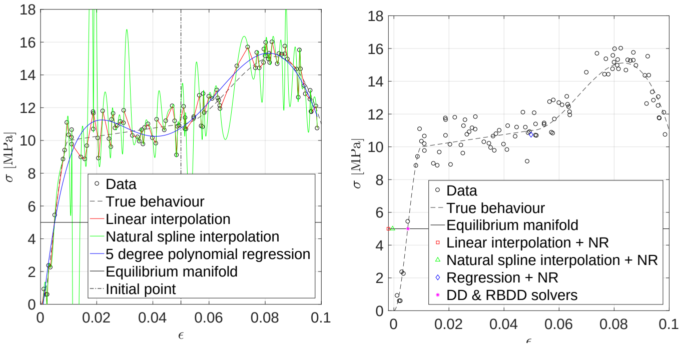
We evaluate the performance of different data-driven solvers, including the reliability-based one proposed herein. As it could be predicted, the main problem of the linearization approach appears when dealing with irregular (non-smooth) empirical manifolds. This is typical in Physics when working with models that have discontinuities, like in many mechanical problems such as plasticity, damage, fracture and contact problems. A very basic unidimensional trivial problem exemplifies well their main pathologies.

For a simple uniaxial loaded rod, with $F = 100$ kN, $A = 200$ cm^2 and $L = 10$ m, this problem may be easily solved through traditional model-based techniques. The solution is based on the combination of three equations. Equilibrium equation, $\sigma A = F$, compatibility equation, $\varepsilon = \frac{u}{L}$. For this problem to be mathematically closed, we need a mathematical relation, i.e. a model, relating the internal (state) variable stress, σ, and the measurable variable strain, ε, what is known as a constitutive relation of the material $\sigma = f(\epsilon)$.

Let us consider that the constitutive relation is not known and the material behavior could be linear, smoothly nonlinear or non-smoothly nonlinear. In any case, what we have to describe the material behavior is a considerable amount of experimental pair values (ε, σ), $\mathcal{E} = \{(\varepsilon^j; \sigma^j)\}_{j=1,\cdots,m}$. For testing data-driven solvers based on linearization, let us compare the computed results when considering a non-smoothly nonlinear behavior and using the well-known iterative tangent Newton-Raphson method, with the analytical results obtained through the exact linear model.

Figure 1a shows the considered empirical set, the equilibrium manifold and the constitutive manifold built for some fitting techniques (linear interpolation, natural spline interpolation and 5 degree polynomial regression). The vertical dashed line shows initial points considered for the Newton-Raphson solver. Figure 1b shows the empirical set, the equilibrium manifold and final point for each solver. Both reliability-based data-driven (RBDD) and data-driven (DD) solvers converge to the same point. Even if polynomial regression is not sensitive to noise, convergence is not achieved for classic solvers because of the untrue local convexity of the built manifold. Besides, due to noise, natural splines suffer spurious oscillations provoking bad convergence. This can be avoided using linear interpolation, but in this case, non-smoothness of the broken line is incompatible with a tangent-based solver, which in turns results in non-convergence.

Figure 2 shows solution points for the DD and RBDD solvers for a material where the uncertainty associated with



(a) Built manifold. (b) Solution point.

Fig. 1. Performance of different tested solvers.

the actual material behavior is not homogeneous: in the elastic zone, where the material is very well characterized, uncertainty is low, but it increases when strains are higher. RBDD solver is sensitive to this variation, while DD solver is not. For complete information, Figure 2 is complemented by the statistical properties summarized in Table I.

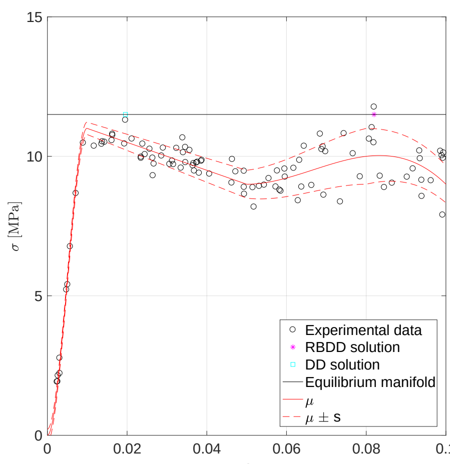


Fig. 2. Performance of DD and RBDD solvers for heterogeneous uncertainty.

TABLE I
STATISTICAL PROPERTIES FOR BOTH SOLVERS.

DD	Squared distance	$3.90 \cdot 10^2$
RBDD	Squared distance	0.08
	Expected value	3.33
	Variance	11.32
	95%-Confident bound	10.06

IV. Conclusion

A new reliability-based data-driven (RBDD) solver has been formulated for data-driven simulation-based engineering problems, allowing uncertainty considerations in the input data that are, therefore, not considered as uncertainty-free, but of random nature. The data-driven simulation problem is defined as a constrained stochastic optimization problem. It has been shown that selecting a proper uncertainty dependent distance, the Mahalanobis distance, results in very good statistical properties as well as easily interpretable optimal distances. RBDD solvers present a meeting point between theoretical sciences, through epistemological constraints, and experimental sciences, through uncertain real world data. The elegance of the mathematical formulation enables many analysis and theoretical considerations for the whole spectrum of Continuum Physics. The ease of combining the presented concepts with all trendy Data Science and Deep Learning tools opens up huge possibilities for facing the most challenging problems.

Acknowledgment

The authors gratefully acknowledge the financial support from the Spanish Ministry of Economy and Competitiveness (MINECO MAT2016-76039-C4-4-R, AEI/FEDER, UE), the Government of Aragon (DGA) and the Biomedical Research Networking Center in Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN). CIBER-BBN is financed by the Instituto de Salud Carlos III with assistance from the European Regional Development Fund.

This work is part of the paper *A new reliability-based data-driven approach for noisy experimental data with physical constraints*, in publication process.

References

[1] S. Hill, F. Provost, and C. Volinsky, “Network-based marketing: Identifying likely adopters via consumer networks,” *Statistical Science*, pp. 256–276, 2006.

[2] C. S. Aneshensel, *Theory-based data analysis for the social sciences*. Sage, 2013.

[3] W. Raghupathi and V. Raghupathi, “Big data analytics in healthcare: promise and potential,” *Health Information Science and Systems*, vol. 2, no. 1, p. 1, 2014.

[4] F. Darema, “Dynamic data driven applications systems: A new paradigm for application simulations and measurements,” in *International Conference on Computational Science*. Springer, 2004, pp. 662–669.

[5] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Journal of basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[6] B. Peherstorfer and K. Willcox, “Dynamic data-driven reduced-order models,” *Computer Methods in Applied Mechanics and Engineering*, vol. 291, pp. 21–41, 2015.

[7] S. Xiao and T. Belytschko, “A bridging domain method for coupling continua with molecular dynamics,” *Computer methods in applied mechanics and engineering*, vol. 193, no. 17, pp. 1645–1669, 2004.

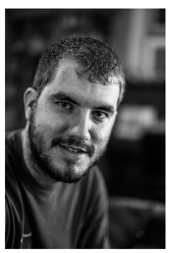
[8] T. Sussman and K.-J. Bathe, “A model of incompressible isotropic hyperelastic material behavior using spline interpolations of tension–compression test data,” *Communications in numerical methods in engineering*, vol. 25, no. 1, pp. 53–63, 2009.

[9] R. Ibanez, E. Abisset-Chavanne, J. V. Aguado, D. Gonzalez, E. Cueto, and F. Chinesta, “A manifold learning approach to data-driven computational elasticity and inelasticity,” *Archives of Computational Methods in Engineering*, pp. 1–11, 2016.

[10] P. Ladevèze, “The large time increment method for the analysis of structures with non-linear behavior described by internal variables,” *COMPTES RENDUS DE L ACADEMIE DES SCIENCES SERIE II*, vol. 309, no. 11, pp. 1095–1099, 1989.

[11] T. Kirchdoerfer and M. Ortiz, “Data-driven computational mechanics,” *Computer Methods in Applied Mechanics and Engineering*, vol. 304, pp. 81–101, 2016.

[12] R. De Maesschalck, D. Jouan-Rimbaud, and D. L. Massart, “The mahalanobis distance,” *Chemometrics and intelligent laboratory systems*, vol. 50, no. 1, pp. 1–18, 2000.



Jacobo Ayensa Jiménez is a Ph.D. student at the Mechanical Engineering Department, University of Zaragoza, Spain. He received his B.Sc. in mathematics and M.Sc. in Civil Engineering at the CFIS center, in the Polytechnic University of Catalonia, Spain, and his M.Sc.in Advanced Design in Mechanical Engineering in the University of Seville, Spain. He began his professional career in the research department of Abengoa working on the uncertainty incorporation in mechanical systems design, such as CSPP troughs and heliostats. Then he moved to the academic world for starting his Ph.D at the Aragón Institut of Engineering Research, where he is focused in mathematical modeling of tumour microenvironements.